Proposed Panel Conclusions and Recommendations for the Isolate Chicken Eye (ICE) Test Method

Expert Panel Meeting

National Institutes of Health
Natcher Conference Center
Bethesda, Maryland
January 11-12, 2005

ICE Test Method Primary Reviewers

- Robert Scala, Ph.D., *Tuscon, AZ*
- Roger Beuerman, Ph.D., *Louisiana State University*
- June Bradlaw, Ph.D., *International Foundation for Ethical Research*
- A. Wallace Hayes, Ph.D., D.A.B.T., F.A.T.S., F.I.Biol, F.A.C.F.E., E.R.T., *Harvard School of Public Health*
- Robert Peiffer, D.V.M., D.A.C.V.O., *Merck Research Laboratories*
- Nancy Flournoy, Ph.D., *University of Missouri* (Biostatistician)

BRD Section 1.0:
ICE Test Method Rationale

1.1 Scientific Basis for the ICE Test Method

- Mechanistic basis not known.
- It is unclear if similar effects occur in the chicken relative to the rabbit (or human).
  - Anatomy and structure of the eye are different.
  - ICE does not include the tear film, and tears are an essential component of normal surface physiology and protection.
- However, ICE test does not necessarily have to be mechanistically based (an accurate correlation to the ocular irritancy classification of a test substance is the primary goal).
- Add discussion of cellular mechanisms of corrosion and severe irritation (e.g., necrosis, apoptosis) and relevance to *in vitro* testing. (ICE, HET-CAM)
- Discuss the role of responsive inflammatory cells in isolated rabbit eyes. (ICE, CAM in HET-CAM) (same thing…)

1.2 Regulatory Rationale and Applicability

- ICE vs. *In Vivo* Rabbit Test:
  - ICE evaluates only corneal effects and does not account for effects on the iris and conjunctiva.
  - ICE does not account for the reversibility of corneal effects.
  - ICE does not account for systemic effects.
  - ICE is a short-term test and may not identify slow-acting irritants.
  - Unlike the *in vivo* rabbit eye test, ICE does not assess iris, conjunctiva, including limbus, or systemic damage

BRD Section 2.0: ICE Test Method Protocol Components

2.1 Test Method Components for the Recommended Version of the Protocol

- Description of protocol components is adequate.
- Areas of concern include:
  - different slit-lamp systems yield varying results
  - temperature is not well controlled
  - the drip system would seem to be difficult to adjust to ensure that the whole cornea is superfused properly
  - a horizontal position of the apparatus would eliminate the need for removal of the eyes during dosing
  - the process of randomization of the eyes is not described
  - the exposure time (10 sec) may be too brief
  - the number of replicates (n = 3) is small
  - the assay medium (isotonic saline) does not contain divalent cations
  - both positive and negative controls should be routinely included
  - guidelines on when histopathology is necessary are not well described
  - Identification of reference substances that are part of the performance standards developed for the validated test method
  - The BRD should identify the decision criteria (Prediction Model) for identifying ocular corrosives and severe irritants

2.2 Basis for Selection of the Test Method System

- No concerns with regard to this section of the BRD.

2.3 Proprietary Components

- No concerns with regard to this section of the BRD.

2.4 Number of Replicate and/or Repeat Experiments for Each Test

- n = 3 eyes per test substance with a single experimental replicate seems limited
  - an adequate assessment of between eye variability would be difficult
- The appropriate number of eyes to be used per test substance is not clear
  - reportedly reduced from 5 to 3 with no adverse effect on ICE performance

2.5 Study Acceptance Criteria

- Positive controls should be included in the acceptance criteria

2.6 Basis for any Modifications to the Original Test Method Protocol

- Only modification has been the reduction of eyes from 5 to 3 per test substance
  - No formal evaluation on the optimum number of eyes
  - The affect of this change on ICE performance is not known

2.7 Adequacy of the Recommended Standardized Protocol

- Only change is the inclusion of a positive control
- Several areas may need optimization:
  - the optimum mechanism to handle differences in corneal swelling values for test substances from different laboratories
  - the capacity of the custom superfusion apparatus (i.e., can it be increased without adversely affecting results?)
  - the assay medium (i.e., inclusion of divalent cations)

BRD Section 3.0: Substances Used for Previous Validation Studies of the ICE Test Method

3.1 Types & Numbers of Substances/Products Used for Prior Validation Studies

- An acceptable range of organic and inorganic substances which encompassed the range of irritancy were evaluated
- The total number of substances evaluated ≥ 92 substances likewise is a reasonable number for validation studies

3.2 Coding Procedures for Test Substances and Quality of ICE Test Method Data

- Only one study (Balls et al. [1995]) made reference to the use of coded substances.
- However, in a retrospective evaluation, lack of coding does not appear to be justification for rejecting the data.

BRD Section 4.0: *In Vivo* Reference Data Used for an Assessment of Test Method Accuracy

4.1 *In Vivo* Rabbit Eye Test Method Protocols Used to Generate Reference Data

- The *in vivo* rabbit eye test method protocol(s) used to generate reference data in the cited studies were appropriate.

4.2 Interpretation of *In Vivo* Test Method Results for Cited Studies

- The interpretation of the results of the *in vivo* rabbit eye tests was correct
- The concern can reasonably be raised that these regulatory classification methods may be less than adequate for use in evaluating or making distinctions between *in vitro* methods and their suitability for chemical or product class evaluations.
- However, the background *in vivo* data for some of the substances were available for NICEATM review

4.3 Data Quality for Test Substances When Original Study Records Are/Are Not Available

- Original study records were not available for any of the reports evaluated.
- However, an evaluation of the results could be made and the quality of the studies otherwise appears to be adequate.
- Future validation studies should be conducted under GLP compliance and original study records should be readily available

4.4 Data Quality With Respect to Extent of GLP Compliance

- Not all studies evaluated for the ICE test method reference GLP compliance
- However, a distinction in the weight given to GLP-compliant v. non-GLP-compliant studies in the BRD may not be necessary
  - GLP regulations deal more with documentation than with the actual performance of the tests documentation
  - When the basic requirements of the GLP procedure (the "spirit" of GLPs) have been implemented in a study, lack of complete/formal GLP compliance is not an adequate criterion to exclude *in vivo* or *in vitro* data from the evaluation of the performance of a toxicity test

4.5 Availability of Relevant Human Ocular Toxicity Information

- There needs to be greater effort to obtain and consider information on human topical ocular chemical injury.
- Most of the available information originates from accidental exposure

4.6 Accuracy and Reliability of the *In Vivo* Rabbit Eye Test

- The potential variability of the *in vivo* rabbit data has not been adequately discussed.
- Evaluation of an alternative method is unavoidably biased by the selection of the *in vivo* data used in that evaluation
- Any optimization and validation studies should use existing animal data; if available.
- Additional animal studies should only be conducted if important data gaps are identified and such studies should be carefully designed to maximize the amount of pathophysiological information obtained (e.g., wound healing)
- Minority opinion – no animal testing for this purpose

BRD Section 5.0: ICE Test Method Data and Results

5.1 ICE Test Method Protocols Used to Generate Each Set of Data

- The ICE protocols used in each of the published validation studies are adequately described.

5.2 Other Comparative ICE - *In Vivo* Rabbit Eye Test Data Not Considered in the BRD

- Additional comparative ICE - *in vivo* data do not appear to be available.

5.3 Statistical and Nonstatistical Approaches Used to Evaluate the Resulting ICE Data

- The approaches used to evaluate the ICE data appear to adequately describe the accuracy and reliability of the test method.
- However, given the unavailability of original ICE data, a definitive statement regarding the adequacy of these approaches is not feasible.

5.4 Use of Coded Substances, Blind Studies, and GLP Guidelines for Cited Studies

- Coding of test substances was carried out in only one study
- However, the absence of coding is not adequate justification for rejecting the data from these studies.

5.5 "Lot-to-Lot" Consistency of Test Substances and Timeframe of Studies

- The test substances and the concentrations used were adequately described in the BRD.

BRD Section 6.0: ICE Test Method Accuracy

a) The closeness of agreement between a test method result and an accepted reference value.

b) The proportion of correct outcomes of a test method

6.1 Accuracy Evaluation of the ICE Test Method for Identifying Ocular Corrosives and Severe Irritants as Defined by the EPA (1996), the EU (2001), and the GHS (2003)

- The method appears to perform equally well for the three
- classification systems.
  - Similarities likewise occur in discordant substances.
- The overall false positive rate (8-10%) seems adequate.
- However, the acceptability of the false negative rate (30- 40%) is less evident since this would result in corrosives/severe irritants to be tested *in vivo* (according to the tiered testing strategy)
- A comprehensive accuracy assessment in the absence of suitable human data should take account of the variability in the Draize test itself.
- A reliability analysis of the Draize test (as is referenced in the BRD) would be useful
- In addition to the analyses conducted, the Panel suggests an assessment based on ranking of experimental data for severity for both the reference method and the in vitro test

6.2 Strengths & Limitations of the Test Method, Including Those Applicable to Specific Chemical Classes or to Certain Physicochemical Properties

- Discordant results in the ICE test relative to the *in vivo* classification most often were attributed to either surfactants (4/7 false negative) or alcohols (5/10 false positive).
- Limitations with respect to solids and insoluble substances, which may be underpredicted, are also adequately described.
- Use of ICE in a tier-testing strategy creates the potential for exposure of animals to corrosives due to the excessive false negative rate.

6.3 Issues of Data Interpretation

- **The endpoints used are adequately described.**

BRD Section 7.0:
ICE Test Method Reliability (Repeatability/Reproducibility)

A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time.

7.1 Selection Rationale for the Substances Used to Evaluate Test Method Reliability

- Only one study was used for the evaluation of reliability.
- The selection rationale for substances used in this study were adequately described.
- Substances that were tested covered a broad range of products and individual chemicals
- Both solid and liquid materials and polar and non-polar materials.
- The full range of irritancy potential was represented.

7.2 Analyses & Conclusions Regarding Intralaboratory Repeatability and Intra- & Inter-laboratory Reproducibility

- Test method reliability analyses and conclusions are sound and appropriate.
- Both qualitative and quantitative evaluations of interlaboratory variability were conducted appropriately.
- No intralaboratory repeatability and reproducibility were conducted because of a lack of appropriate information.

7.3 Availability of Historical Negative & Positive Control Data

- Historical negative and positive control data were not available.

7.4 Effect of Minor Protocol Changes to Recommended Test Method Protocol and Transferability of Test Method

- The recommended version of the *in vitro* ICE test method will, like most *in vitro* tests, be somewhat sensitive to protocol changes.
- Any validation study of this test should use a standard test protocol that is not altered by the testers.

BRD Section 8.0: ICE Test Method Data Quality

8.1 Extent of Adherence to GLP Guidelines and Use of Coded Chemicals

- The extent of adherence to national and international GLP guidelines for the three studies reported in the BRD is not adequately presented.
- Some studies do not state in a definitive manner that they were conducted under GLP.
- Without assurance of GLP guidance including sample coding, the quality of the data cannot be easily verified.

8.2 Data Quality Audits

- No information regarding data quality audits was reported for any of the three ICE studies.
- The original data for each of the ICE test method experiments was not reviewed (the BRD states that such data were not readily available).
- While data quality audits may not be feasible for the retrospective data, original data should be readily available for examination for any future validation studies.

8.3 Impact of Deviations from GLP Guidelines

- Information on GLP deviations or their impact on the study results is reportedly not available.
- In the absence of such information, the validation status of the ICE may be questioned.

8.4 Availability of Laboratory Notebooks or Other Records for an Independent Audit

- The lack of available laboratory notebooks or other records of the raw data has been addressed.
- No raw data were used in these evaluations
- Although caution should be exercised when evaluating these data, the lack of original records does not appear to be a rationale for excluding these data.
  - Access to the original *in vitro* and *in vivo* data would allow for a more complete retrospective evaluation of ICE.
- However, any future validation studies should include coded test substances of known purity, from a common source and centrally distributed; appropriate controls and conducted under GLP guidelines.

BRD Section 9.0: Other Scientific Reports and Reviews

9.1 Adequacy and Completeness of Relevant Data Identified in Other Published or Unpublished ICE Studies

- Information/data from two additional sources
- Inadequate information on the substances tested (identity not specific) and/or on the results obtained from the *in vitro* or *in vivo* studies precluded an assessment of the performance characteristics of the ICE

9.2 Adequacy and Completeness of the Conclusions Published in Independent Peer Reviewed Reports or Other Independent Scientific Reviews

- The conclusions have been adequately discussed and compared

9.3 Approaches that can be Used to Expedite the Process for Obtaining Additional In-House Data from the Private Sector

- The use of the Federal Register requesting information was not productive (only one response)
- Personal contacts by the agencies to which data have been submitted may be the best method to secure additional in-house data from the private sector
- If such data are not received, additional *in vivo* studies may be necessary to compile an adequate reference database

BRD Section 10.0: Animal Welfare Considerations (Refinement, Reduction, Replacement)

10.1 Extent to Which the Test Method Will Refine, Reduce or Replace Animal Use

- While the ICE test both refines and reduces animal use, it is probably best characterized as a partial replacement under the 3Rs of refinement, reduction, and replacement
- There is no additional inflicting of pain or distress to the animal as a result of the testing procedures
- Because chickens do not come under the Animal Protection Act, there is still a need to ensure their humane treatment.

BRD Section 11.0: Practical Considerations

11.1 Adequacy and Completeness of Test Method Transferability

- Facilities and major fixed equipment, along with other necessary equipment are adequately described.
- Transferability of the test method does not appear to be a significant obstacle to its use.
  - However, specifications for the custom built superfusion apparatus must be readily available
- A training video and other visual media on the technical aspects of the assay is recommended (place in all)
- Training approaches in the application of this test method should be developed/implemented (place in all)

11.2 Adequacy and Completeness of Test Method Training

- The training required to conduct the ICE test is dependent on the background and experience of the person.
- ICE requires:
  - good manual de xterity
  - knowledge of the anatomy of the eye
  - the ability to recognize an unacceptable specimen
  - accurate evaluation of the results at the requisite time points
  - familiarity with using of a slit-lamp

11.3 Adequacy and Completeness of the Information on the Cost Involved in Conducting a Study Using the ICE Test Method and Comparison to the Cost of Conducting the *In Vivo* Rabbit Eye Irritation Test

- It would appear that the cost of conducting the ICE with all the necessary controls and in triplicate is comparable with the cost of conducting a 3 day/3 animal study

11.4 Adequacy and Completeness of the Information on the Amount of Time Needed to Conduct a Study Using the ICE Test Method and Comparison to the Time it Takes to Conduct an *In Vivo* Rabbit Eye Irritation Study

- The ICE test (6-8 hr) would significantly reduce the time needed to assess the ability of a test substance to induce ocular corrosivity or severe irritancy relative to the *in vivo* rabbit eye test (typically at least 3 days and may extend up to 21 days). However, it is recognized that a corrosive or severe irritant may be detected within a few hours using a single rabbit.

BRD Section 12.0: Proposed ICE Test Method Recommendations

12.1 Recommended Version of the ICE Test Method

- The ICE protocol has changed only slightly since its development.
- With regards to the recommended protocol:
  - it is unclear if the appropriate number of eyes are being used to ensure optimum performance. Therefore, an evaluation of existing data using 5 eyes or 3 eyes would be useful
- It is unclear if the use of maximum mean scores is the most appropriate scoring system to ensure optimum performance
- The low overall false positive rate means that the ICE test can be used at present to screen for severe eye irritants.
  - However, given the high false positive rates calculated for a small number of alcohols, caution should be observed when evaluating ICE test results with these types of substances.

12.2 Recommended Standardized ICE Test Method Protocol (I)

- The ICE test method appears to be useful in the identification of ocular corrosives/severe irritants in a tiered testing strategy, with the following limitations:
  - Alcohols tend to be overpredicted
  - Surfactants tend to be underpredicted
  - Solids and insoluble substances may be problematic as they may not come in adequate contact with the corneal surface (leading to underprediction)
- Users should be aware of the risk of BSE and other zoonoses and use proper precautions

12.2 Recommended Standardized ICE Test Method Protocol (II)

Not withstanding what is stated in 12.1

- The reliability of the ICE test method has not been adequately evaluated.
- The appropriateness of using only three eyes per test substance has not been formally evaluated
- The method of dosing the eyes (i.e. removal from the superfusion apparatus and turning horizontally) would seem more effective if the eyes remained in a horizontal position throughout the assay
- A standardized scoring scheme for histopathology should be defined using the formal language of pathology to describe any effects
- The appropriate circumstances under which histopathology would be warranted should be more clearly defined.
- To maximize the likelihood of obtaining reproducible results, reference photographs for all subjective endpoints (i.e., corneal opacity, fluorescein retention, and histopathology) should be readily available.

12.3.1 Recommended ICE Optimization Studies (1)

- **Additional studies are recommended to:**
  - optimize the ICE decision criteria with respect to each regulatory classification scheme in order to reduce the false negative rate without unacceptably increasing the false positive rate
  - determine the most appropriate endpoint score to be used for categorization (i.e., a multivariate analysis)
  - determine the impact on performance of routinely repeating experiments to ensure consistency
  - determine the appropriate number of eyes to be tested to ensure optimum ICE performance
  - Evaluate the impact of delayed use of chicken eyes on assay performance
  - Centering lights should be installed on the optical pachymeter to ensure consistent central corneal thickness measurements across labs

12.3.1 Recommended ICE Optimization Studies (2)

- **Additional studies are recommended to:**
  - determine the most appropriate use of histopathology as an endpoint (i.e., precisely when it should be used)
  - determine the optimum mechanism for handling differences in corneal swelling values for test substances from different laboratories (I.e, correction factor, or something more appropriate)
  - the impact of expanding the capacity of the custom superfusion apparatus (i.e., can it be increased without adversely affecting results?)
  - determine the most appropriate assay medium (i.e., inclusion of divalent cations)
- **Reference substances should be identified that can be used as part of performance standards**
- **NICEATM/ICCVAM should facilitate the development of a histopathology scoring system for corneal damage (with visual aids)**

12.3.2 Recommended ICE Validation Studies

- Any optimization and validation studies should use existing animal data; if available.
- Additional animal studies should only be conducted if important data gaps are identified and such studies should be carefully designed to maximize the amount of pathophysiological information obtained (e.g., wound healing)
- Minority opinion – there is sufficient data so no animal testing for this purpose
- NICEATM/ICCVAM should facilitate the development of a histopathology scoring system for corneal damage (with visual aids)